DOCUMENT RESUME

ED 210 650 CS 006 401

AUTHOR Marschner, Jane; And Others
TITLE Assessment of the Cognitive Domain of Reading.
PUB DATE Mar 81
NOTE 50p.; Paper presented at the Annual Meeting of the Missouri State Council of the International Reading Association (14th, Columbia, MO, March 13-14, 1981).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Achievement Tests; *Diagnostic Tests; Educational Assessment; Elementary Secondary Education; *Reading Readiness Tests; *Reading Tests; *Test Reviews; *Test Selection

ABSTRACT

Noting that the purpose of testing and test validity, reliability, norms, and usefulness should be primary considerations in choosing a reading test, this paper is intended to serve as a guide in test selection. The document provides a checklist for collecting the information necessary for selecting a test, and brief overviews of five reading achievement/survey tests, nine diagnostic tests, and four reading readiness tests. A chart for quick reference follows each set of descriptions. The document concludes that a comprehensive testing program should include the use of teacher observation, health and anecdotal records, work samples, classroom tests, cumulative records, case studies by specialized consultants, objective or standardized tests, and any other tests needed to assess student performance. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED210650

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Assessment of the Cognitive Domain of Reading

Jane Marschner
Mary Spickler
Ann Stahl

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Mary Spickler

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS 006 401

ASSESSMENT OF THE COGNITIVE DOMAIN OF READING

This chapter is an examination and evaluation of a variety of instruments designed to assess reading performance. We hope to help the school answer the question: "Which reading test should we use?" At first glance, test selection appears to be a routine procedure that could be handled with minimum preparation. However, the wise selection of reading tests calls for both considerable understanding and knowledge of various reading components.

Reading assessment specifically seeks to know how well a student presently reads and what specific skills he/she may need to know in order to improve reading performance. It should help identify children in need of special assistance. In order to gain this information, different types of tests should be utilized. A primary consideration in choosing a test should be an examination of the purpose of testing. According to Schreiner, it is best not to test at all if the uses of the test results are not known in advance. (Schreiner, 1979)

There is no single test that will be "best" for all pupils, all curriculum objectives, all purposes, and all uses. For that reason, a variety of test types will be examined and evaluated. First, standardized norm referenced tests will be studied. These include survey and/or achievement tests, diagnostic tests, and readiness tests. Then information will be presented on the various criterion referenced tests: informal

tests, diagnostic tests, and those that are objective based. It is hoped that this information will enable the reader to become more familiar with the strengths, weaknesses, and purposes of particular tests, thereby allowing selection of the best instrument available for the specified purpose. (Schreiner, 1979)

Test selection should include an examination of several major factors: validity, reliability, norms and usability.

Validity, the most important characteristic of any test, describes the degree of accuracy with which a test measures what it is intended to measure. Factors such as the various pupil responses, nature of the group, the test administration and scoring, and factors within the test itself (unclear directions, poorly constructed items, inappropriate level of difficulty, test too short, etc.) can all adversely affect validity.

There are several types of test validity. Face validity is the general appearance of the test; the term describes the test validity from the layman's point of view. Content validity requires subject matter expertise. A thorough examination of the test items to determine if it contains a representative sample of tasks covering the objectives in a curriculum is necessary. (Martuze, 1977) Construct validity refers to the extent to which the scores are meaningful in light of our theories. Determination of construct validity is mady by: (1) identifying the constructs presumed to account for test per-

formance, (2) derive hypotheses regarding test performance from the theory underlying the construct, and (3) verify the hypotheses by logical and empirical means. (Gronlund, 1976) Criterion referenced validity involves comparing test scores with some other measurement of the traits being assessed. Obviously, this measure is only good if the other measure is known to be a valid measure itself. Another problem with this method of determining validity is that alghough the number of items missed by students on both tests may correlate, the students may miss different items on each of the tests. This makes little difference on a survey test, but is very misleading on a diagnostic or criterion-referenced test that purports to diagnose specific blends, digraphs, etc. (Ekwall, 1976) If the test is administered about the same time as the other measurement, we say that we are estimating the concurrent validity. If the test is administered, and then after a substantial period of time the other measurement is used, we are estimating the test's predictive validity. The degree of criterion related validity is expressed with a correlation coefficient. A coefficient close to one indicates high validity; close to zero, a low validity. (Schreiner, 1979)

Reliability is a term used to describe the accuracy of test measurement; i.e., Does the test give consistent results? Factors such as examinee differences, the scores, the testing situation, time influences, and test content can have an effect on test reliability. (Schreiner, 1979) Reliability is usually

stated as a correlation coefficient or as a standard error of measurement. The higher the coefficient or the lower the standard of measurement, the more reliable the test. (Harris, 1975)

Several methods of determining reliability are commonly used. The test-retest method involves administering the test and then repeating it at a later date. This accounts for time errors. The split-half method involves administering the test once, and then getting two scores by using the odd numbered items for one score and even numbered items for the other score for comparison. This method measures content errors. The equivalent form method requires that two tests are constructed with similar items and administered one right after the other for comparison of scores. This method of determining reliability is through the Kuder-Richardson formula (KR-20). One test is administered, and the results divided into halves. A statistical formula is then applied based on the number of correct responses. This method is inappropriate for tests in which speed is an essential factor because if all items were not attempted the correlation coefficient will be inflated. (Wallace & Larsen, 1978)

Norm information is very important when test results are to be used for comparisons of individuals. The term refers to the statistical information which describes the distribution of scores of the population of examinee's on a particular test. Grade norms are based on the grade group in which the student's

raw score is average. Age norms are the age group in which the student's raw score is average. Percentile norms is the percentage of students in the reference group who fell below the student's raw score. Standard score norms is the distance of a student's raw score above or below the mean of the reference group in terms of standard deviation units. (Wallace & Larsen, 1978)

Even if the reliability and validity credentials of a test are impeccable, referencing a test score to an inappropriate norm population will result in an invalid interpretaion of that score. In evaluating normative information, consideration must be given to the definition of the norm group, the procedure used to construct the norm sample, and the usability of the norms. (Martuza, 1977)

Usability is another concern in test selection. This determines whether a test is practical for individual use. It includes such factors as ease of administration, time required for administration, ease of scoring, ease of interpretation and application, availability of equivalent forms, adequacy of the test manual, and cost. (Martuza, 1977)

The following checklist developed by Kavale (cited in Schreiner, p. 15) is included to serve as a guide in collecting the information necessary for selecting a test.

Test Selection Checklist

I. What is your purpose for evaluating this test?
   A. Why do I need a test?
   B. What information do I require?

II. Test Overview
   A. Test information
      1. Name of test - what does name imply about purpose?
      2. Author
      3. Publisher
      4. Date of publication and last revision
   B. Target population
      1. Age level - appropriate for your situation?
      2. Grade level - appropriate for your situation?
      3. Individual or group administration
      4. Number of available forms
   C. What construct or quality is this test designed to measure?

III. Practical features
   A. Time considerations
      1. Time required to administer total test
      2. Could the time requirement be handled in your situation?
      3. Time required to score - how involved and time-consuming is the scoring procedure?
      4. Who could score the test in your situation? Does scoring involve judgment or simply tallying right and wrong responses?
      5. Is machine scoring available?
   B. Financial considerations
      1. Cost of test booklet
      2. Cost of answer sheets
      3. Cost of test manuals
      4. Cost of replacement materials
      5. Cost of scoring services
   C. General considerations
      1. Are the directions for administering clear and is the language appropriate for the students?
      2. Are the directions for scoring clear?
      3. Are answer sheets appropriate for age and ability level of students?
      4. What training is required to administer the test?
   D. Format design
      1. Are illustrations current?
      2. Is print size adequate?

IV. Validity
- A. What evidence of validity is given?
- B. Does reported validity appear adequate in relation to intended purposes?
- C. Subtest validity
  1. Does the name adequately describe the subject?
  2. Does the subtest seem logically related to the test as a whole?

V. Reliability
- A. What evidence for reliability is given?
- B. Do the reported reliabilities appear adequate?
- C. Subtest reliability
  1. Are reported coefficients adequate?
  2. Are subtests sufficiently long?

VI. Norms (when appropriate)
- A. Are the norms provided adequate and appropriate?
- B. Does the normative sample seem truly representative?
  1. Up-to-date?
  2. Age--does the sample cover age range in your situation?
  3. Sex--were appropriate numbers of boys and girls included?
  4. Socioeconomic level--does sample cover SES found in your situation?
  5. Range of ability--does the sample include ability levels in your situation?
- C. Are the norms relevant for your group and purposes?

VII. Bias
- A. Is there evidence in the test content of sex stereotyping?
- B. Is there evidence in the test content of racial or ethnic bias?

VIII. General evaluation
- A. What is your general evaluation of the test?
- B. For what purpose and group would you recommend this test?

Types of Available Tests

A norm referenced test is one that is tried out with a sample of examinees called a norm group so that an individual's performance can be described relative to that group. Often, the norm group consists of a selected sample drawn from across the nation. (Schreiner, 1979) The advantage of a norm referenced test is that it allows comparison of an individual's score (or class, school, or district scores) to a much larger group of students (the norm group). However, a description of this norming population is critical for interpretation of scores. (Farr & Anastasiow, 1975) If the test has been normed with a large population of bright students, then scores for an average child may be lower than his current academic status; the reverse would also be true. It is important to consider the three "R's" when choosing a norm referenced test. The norms should be: Recent - preferably within the last five years; Relevant - they should have the same characteristics as the group being tested; Representative - they should be a sample of all individuals taking the test. Alert test users should, therefore, look for norm groups that meet these criteria or collect scores on their own pupils and use local norms. (Sage, 1967)

A standardized test is constructed of items that are "appropriate in difficulty and discriminating power for the intended examinees." (Spache, p. 169) This test has uniform prescribed procedures for administration and scoring. It has

specific directions that must be given, time limits that must be kept, and the scoring must follow definite rules. A standardized test also has accompanying tables of norms. One advantage of standardized tests is that they contain concise, uniform directions that ensure that the results are not depressed or inflated because of unclear directions. (Farr & Anastasiow, 1975) Another advantage is that the test scores are usually reported in a variety of ways; raw score, percentile, grade equivalents, and stanines. Also, parallel forms increase the probability of obtaining comparable test scores with repeated testing. Every standardized test is accompanied by a manual which should be written in a clear, concise manner, easily understandable to the teacher administering the test. This manual should contain the norm population, and the procedures the authors used to determine reliability and validity. This type of test is usually easy to administer and has greater reliability than teacher made tests. (Spache, 1976)

A major limitation of standardized tests lies not with the test itself, but with the interpretation of its results... attributing more meaning than they deserve. The test scores may mislead naive or inexperienced test interpreters. It must be remembered that the scores are not fixed measured, they are estimates. (NEA, 1977) The most common error revolves around the grade equivalent score. This score is commonly interpreted as indicating the grade level of material that a student can

adequately handle. Actually, it merely shows that a student correctly answered the same number of items as others at that grade score. (Spache, 1976) For example, a third grader receives a score of 5.2 on a third grade test. This means that he/she had the same number of correct items as a student in the fifth grade-second month taking the same test. A way to prevent this problem is to stress the reporting of scores in stanines, or ranges that take into account the standard error of measurement, rather than give one number which may be taken as an "absolute."

Other criticisms of standardized tests include: (1) the test content may not reflect local instructional objectives or specific curriculums, (2) the test may be culturally and/or linguistically biased, (3) the test results are too broad and general to give teachers any help in selecting appropriate teaching methods, (4) subtest reliabilities are too low to be used for diagnostic purposes, (5) because these tests are group administered, they may not result in a "true" score for some student types. (NEA, 1977)

A school normally assesses the reading performance of its entire population at given intervals. The most common type of test used for this is a standardized norm referenced Reading Achievement and/or Survey test. These tests are defined as those which provide an approximate level at which a child is functioning in relation to a specified group. (Schreiner, 1979) These tests are usually group administered,

and measure vocabulary, comprehension, and perhaps study skills and rate of reading. They usually give a general achievement score which reflects overall reading performance, as well as specific scores for the subtests. Reading achievement/survey tests are primarily used to measure group or individual progress over time, to assess the general reading performance of a student or group, to determine general areas of strengths and weaknesses, to help identify those who are experiencing reading difficulties, to aid in the selection of appropriate materials, to determine the types of classroom organization, to measure the effectiveness of instruction, and to evaluate new programs or methods. (Harker, 1977)

While standardized norm referenced reading achievement tests are useful in providing data to help teachers, they have their limitations as well. First, they do not supply formal diagnostic information. The quality of information for an individual may be less specific than that obtained for a group of students. These tests should be used primarily as screening devices and caution should be exercised in using the scores when planning for individuals. (Schreiner, 1979) Second, scores may be inflated due to guessing, and tend to represent a frustrational rather than instructional level of reading. Third, the very fact that they are "standardized" may make it inappropriate for use with certain individuals or groups. Fourth, tests at upper grade levels assume ability at lower levels. Therefore, the percentile score is more

important in interpretation than the grade equivalent score which could be grossly inflated. Fifth, group administration may work to the disadvantage of certain individuals. (MacGinitie, 1973) Last, while the test may have excellent content validity nationwide, it must closely match the curriculum content of an individual school, or it will provide misleading information.

The guidelines for selection of a reading achievement/survey test consists of three phases:

1. A study of school characteristics and testing needs. What information is desired?
2. An examination and evaluation of available tests. Does this instrument give the needed information?
3. An in-depth study of a particular test which included:
   (a) Matching the school population with the norm group.
   (b) Matching the contents of the test to curriculum content and objectives.
   (c) Matching the validity, reliability, scoring, and interpretive material of the test with the testing purposes and the prospective uses of the test scores. (Katz, 1973)

Content validity is significant in reading achievement tests and should be examined. Also reliability coefficients of .80 and above are desirable. (Schreiner, 1979, p. 80) The standard error of measurement which "allows for" random

errors that result from such factors as guessing should be low. (Schreiner, 1979)

Major Reading Achievement/Survey Tests

Every May thousands of hours are diverted from actual instruction and channeled to the taking of tests. This practice reflects a major purpose of testing - to find out how much a student has learned in a given year. Administering an achievement test late in the year is necessary to evaluate the instructional program. However, if an overview of a student's current status is desired, then the administration of an achievement test early in the year is beneficial. In most cases, the subtests of reading achievement tests do not seem to be a valid measure of subskills and only total test scores should be used. To determine specific strengths and weaknesses, a diagnostic test should be selected. (These will be discussed later.)

A brief overview of five of the many reading achievement/survey tests available follows. These tests were selected for review based upon their popularity, reputation, and recommendations for use contained in Buros Mental Measurement Yearbook, coupled with journal articles and textbooks on the subject of reading assessment. A chart is placed at the end of the descriptions for quick reference.

Gates-MacGinitie Reading Tests (Second Edition)

Author: Walter H. MacGinitie
Date of Publication: 1978
Publisher: Houghton-Mifflin
Range: Grades 1 through 12
Levels: 7
Subtests: Vocabulary, Comprehension - also gives a total score
Time: 55 minutes
Norms: Adequate - representative proportions of Blacks and Hispanics. Out of level norms may be useful in evaluating remedial programs.
Reliability: Complete data presented. Internal consistency (KR20) as well as alternate forms. Coefficients range from .88 to .95.
Validity: The author acknowledges that validity is dependent upon the match between the goals of instruction and the content skills covered in the test, and provides a series of questions to guide the test user in determining the validity of the test.
Manual: Concise, easy to use.
Summary: The Gates-MacGinitie Reading Tests are a series of group administered tests of general reading achievement. There are 7 levels which permit continuous assessment of reading achievement from the first through the twelfth grade. Each level provides independent scores for the two subtests of comprehension and vocabulary, as well as a total score. The provision for out of level testing may be useful. The short time required for administration is another advantage. These tests appear to be carefully developed and standardized and could be used to estimate general reading achievement, identify those who may need further diagnostic testing, assess growth in general reading ability, aid in grouping students, and perhaps cautiously help in diagnosing subskill deficiencies. (Jongsma, 1980)

California Achievement Tests - Reading (CAT)

Authors: Ernest Tiegs, Willis Clark
Date of Publication: 1970
Publisher: CTB/McGraw Hill
Range: Grades 1 through 12
Levels: 5
Subtests: Vocabulary, Comprehension - also gives total score
Time: 40-50 minutes

CAT (Cont.)

Norms: Questionable. Selected sample was solicited by invitation and 60% of the public school districts and 90% of the Catholic school districts agreed to participate.
Reliability: Adequate for the Reading Subtest. KR-20 range from .93 to .97.
Validity: Questionable. There was no statistical study of validity contained in the Technical Bulletin. Content validity is discussed and the steps in item selection are described.
Manual: Adequate
Summary: The CAT are a nationally standardized group series of achievement tests assessing skill development in the content areas of reading, mathematics, and language arts. The reading test can be used independently and could be considered if seeking a short, well constructed survey test of reading achievement. These tests permit assessment from the first through twelfth grades. Since little data was available concerning the test validity, it would be necessary to inspect the content of the test to determine how closely it matches instruction. It should be noted that reference skills are categorized as part of reading comprehension. This test is easy to administer and use.

Metropolitan Achievement Tests: Reading (MAT)

Authors: W. Curost, H. Bixler, J. Wrightstone, G. Prescott, I. Balow
Date of Publication: 1973
Publisher: Harcourt, Brace, Jovanovich
Range: K.7 through 9.5
Levels: 6
Subtests: Word knowledge (vocabulary), word analysis (only at primary levels), reading (comprehension) - also gives a total score
Time: 40-65 minutes
Norms: Inadequate sample. The number of students is not specified.
Reliability: Split-half reliabilities range from .88 to .94.
Validity: Questionable. No validity data was included in the manual, but is reportedly available in a series of technical supplements. The authors contend that because each school has different curricula, content validity will have to be

judged by the individual school. (Salvia, Ysseldyke, 1978) However, there are no suggested methods on how this is to be accomplished.

Manual: Adequate

Summary: The Metropolitan Achievement Tests are a norm referenced group administered test battery which includes a range of levels of reading tests designed to measure general reading achievement. They would be useful in giving a rough indication of broad areas of strengths and weaknesses in reading. Since a school must match the norm group with its student population, a weakness in this area is significant. Another weakness is the fact that the test range stops at 9.5. This prohibits the continuous assessment of an individual or group with the same instrument through the twelfth grade. The authors have tried to eliminate any racial or sex bias by having items reviewed by members of minority groups and then revising or eliminating items they found offensive. (Buros, 1978)

The Stanford Achievement Tests: Reading Test

Authors: R. Madden, E. Gardner, H. Rudman, B, Karlsen, J. Merwin

Date of Publication: 1973

Publisher: Harcourt, Brace, Jovanovich

Range: 1.5 through 12

Levels: 8

Subtests: Vocabulary (1.5-9.5), reading comprehension, word study skills (1.5-6.9), listening comprehension (1.5-6.9); also includes total score

Time: 40-90 minutes

Norms: Adequately constructed and described.

Reliability: Adequately determined by split-half estimates and KR-20 coefficients range from .85 to .95.

Validity: Adequately tested and described

Manual: Clear, comprehensive, consists of 5 parts and covers such information as test administration, norms, interpretation for both teachers and administrators and technical date. There are two separate manuals. One groups items by instructional objectives; the other described the behaviors samples by each item on each subtest at each level of the SAT, and also provides instructional objectives and suggestions on ways to teach those objectives. (Salvia, Ysseldyke, 1978)

Stanford (cont.)

Summary: The SAT-Reading is part of a norm referenced group administered battery of achievement tests designed to measure reading performance from first through twelfth grades. The SAT actually encompasses grades 1.5-9.5, and then the Survey Test of Academic Skills (TASK) is used to assess high school students. There is a special TASK Level II available for use in community colleges. The manual for the Reading test is clear, comprehensive, and practical. Of particular interest may be the subtest designed to measure listening comprehension which adds another dimension to the testing of reading skills. It is important to remember that although this test may serve a limited diagnostic purpose, they were designed primarily as survey instruments and should be used accordingly. This test seems to be one of the better tests and "is a model of what adequately developed achievement tests should be." (Salvia, Ysseldyke, 1978)

The Comprehensive Test of Basic Skills (CTBS): Reading Expanded Edition

Date of Publication: 1976
Publisher: CTB/McGraw-Hill
Range: K.6 through 12.9
Levels: 6
Subtests: Vocabulary, comprehension, reference skills, and total score
Time: 90-105 minutes, usually in two sessions
Norms: Adequate. Minority groups are represented.
Reliability: Adequate
Validity: Questionable. Little validity data was included in the manual. As with all achievement tests, the primary concern is with content validity and that must be determined by those using the test.
Manual: Very good. Written in a clear style that is easy to understand. Even though there is no specific manual for the Reading tests, it is not difficult for a teacher to administer the reading test using the battery manual.
Summary: The CTBS is a standardized achievement battery of which the reading tests are a major part, They are designed to measure reading achievement from K.6 through 12.9. While this appears to be a well researched test, that is both usable and useful, it is weak in the area of

CTBS (cont.)

content validity. Also, the length and the fact that it usually takes two sessions may be a disadvantage in certain situations. However, if the potential user is satisfied with the content covered, then the CTBS could be a part of the reading program.

Name of Test	Date of Publication	Range	Levels	Subtests	Time	Reliability	Validity	Norms	Manual
Gates-MacGinitie Reading Tests (2nd ed.)	1978	1-12	7	Vocabulary Comprehension	55	Adequate	To be determined by user	Adequate	Good
California Achievement Test - Reading	1970	1-12	5	Vocabulary Comprehension	40-50	Adequate	Question-able	Question-able	Easy to use
Stanford Achievement Tests - Reading	1973	1-12	8	Comprehension Vocabulary Word Study Skills	75-120	Adequate	Adequate	Adequate	Excellent
Metropolitan Achievement Tests - Reading	1973	K.7-9.5	6	Word Knowl. Word Analysis	40-65	Adequate	Question-able; Determined by user	Inadequate	Clear
Comprehensive Test of Basic Skills: Reading Expanded Edition	1976	K.7-12.9	6	Vocabulary Comprehension Reference Skills	90-105	Adequate	Question-able	Adequate	Adequate

DIAGNOSTIC TESTS

A diagnostic reading test is an instrument used to evaluate reading and reading related behaviors. These tests are usually administered individually, but some can be given to a group. The diagnostic tests usually "have a larger number of part scores and corresponding larger number of test items measuring each function than survey tests" (Gronlund, 1976, p. 307) "If the number of subtests is very large, the separate parts may be too short to be accurate measures, even though the total score may be highly reliable." (Harris and Sipay, 1975, pp. 180-1) The "...items on these tests are based on detailed analysis of the specific skills involved in successful reading performance and a study of the most common errors made by students. (Diagnostic) test difficulty tends to be low. The lower difficulty is necessary in order to provide adequate discrimination among those students with ... disabilities" (Gronlund, 1976, p. 307) Most diagnostic tests are norm-referenced, although we are more concerned with individual differences. There is a large variation in the range of skills measured by different diagnostic tests. The examiner must select a test keeping in mind the specific type of information desired. Harris and Sipay (1975) suggest three criteria as guidelines for reviewing standardized diagnostic reading tests:

1) How does the test measure the component skills, and do the subscores represent areas for providing remedial instruction?

2) Are the subscore reliabilities sufficiently high (Above .90) for individual use?

3) Are the intercorrelations among subtests sufficiently low (below .65) to warrant differential diagnosis? (p. 181)

An advantage of the diagnostic test is its ability to identify skill strengths and weaknesses. This allows teachers to plan appropriate reading programs for the students.

"One of the major limitations of diagnostic reading tests is their relative lack of reliability and absence of empirical evidence for their validity. Thus, it is imperative that the diagnostic assessment of reading skills consist of repeated measures of those skills. By repeated observations, which confirm initial impressions or call them into question, examiners can cut down on the amount of error involved and increase the reliability of their scores." (Salvia, Ysseldyke, 1978, p. 164)

A brief description of nine popular diagnostic tests follows. A chart of them is also included as an overview.

NAME OF TEST / COMMENTS	Group Administered	Individually Administered	Norm Data Available	Reliability Available	Validity Available	Number of Reading Subtests	Date of Test	Range	Time	Number of Forms
Diagnostic Reading Scales - Revised Edition	no	yes	ade-quate	yes	yes	3	1972	gr. 1-6	45 min.	1
Durrell Analysis of Reading Difficulty	no	yes	limited	no	no	4 (11 op-tional)	1955	gr. 1-6	varies	1
Gates-McKillop Diagnostic Test	no	yes	limited	no	no	11	1955	gr. 2-6	45-90 min.	2
Gilmore Oral Reading Test	no	yes	ade-quate	yes	no	3	1968	gr. 1-8	15-20 min.	2
Gray Oral Reading Test	no	yes	limited	yes	yes	3	1967	gr.1-adult	30-60 min.	4
*Peabody Individual Achieve-ment Test	no	yes	excel-lent	yes	yes	2	1970	gr. K-12	15-20 min.	1
Silvaroli Classroom Reading Inventory	no	yes	no	no	no	3	1974	gr. pp-6	15-20 min.	3
Stanford Diagnostic Reading Inventory	yes	yes	excel-lent	yes	yes	varies with level	1976	gr. 1.5 - 13	2-3 hrs.	2
*Wide Range Achieve-ment Test	no	yes	ade-quate	yes	yes	1	1965	gr.5 -adult	20-30 min.	1

*All data refers to reading sections only.

DIAGNOSTIC READING SCALES - REVISED EDITION

Author: George D. Spache
Publisher: McGraw-Hill
Date: 1972
Test Objectives: To provide standardized evaluation of oral and silent reading skills and auditory comprehension, grades 1-6.
Subtests:
1. Word Recognition List - This is a graded pretest to determine proper starting level.
2. Reading Passages - These are graded passages from gr. 1.6 to 8.5.
3. Phonics Test - This contains eight parts which include: consonant sounds, vowel sounds, consonant blends and digraphs, common syllables, blending, letter sounds, initial consonant substitution, and auditory discrimination.

Norm Group: The author used 1,269-2,081 pupils who were drawn from both rural and urban schools in four states.

Reliability:

Word Recognition	Reliability (Kuder-Richardson)
List 1	.96
List 2	.87
List 3	.91

Reading Passages	Reliability (alternate form)
Instructional	.99
Independent	.98
Potential	.99

Validity:
1. Content - Each reading selection was geared toward the range of reading done at that level.
2. Construct - The author offers studies that compare his data with potential levels of true gain after remedial help. His correlations range from .32 to .89.
3. Concurrent - Scores from the Scales were compared with total scores of the California Reading Test. Those yielded correlation coefficients from .63 to .92.

Scoring: Each test yields a grade equivalent score. The reading passages also provide a grade equivalent at the instructional, independent, and potential levels of reading.

Limitations: The manual is hard to follow. The author uses a low criterion for the comprehension level. The test as a whole tends to overestimate student performance. (Buros, 1978, Vol. II)

DURRELL ANALYSIS OF READING DIFFICULTY

Author: Donald D. Durrell
Publisher: Harcourt, Brace, Jovanovich
Date: 1955
Test Objectives: To indicate specific skill-developed strengths and weaknesses, grades 1-6.
Subtests:
1. Oral Reading - Consists of eight paragraphs of increasing difficulty to be read orally and timed; then the child responds to literal comprehension questions.
2. Silent Reading - Contains eight paragraphs comparable in difficulty to the Oral Reading subtest. These are timed and then the examiner tests recall.
3. Listening Comprehension - The examiner reads seven paragraphs aloud and asks comprehension questions.
4. Word Recognition and Word Analysis - Words are presented through a tachistoscope for one-half second. If the child cannot recognize it he is shown the word to analyze in an untimed format.

There are eleven supplementary subtests: Naming letters, Identifying letters named, Matching letters, Visual memory for words, Hearing sounds in words, Learning to hear sounds in words, Sounds of letters, Learning rate, Phonic spelling of words, Spelling, Handwriting.

Norm Group: In the manual Durrell states he used 1,000 children but gives no further information.
Reliability: None given
Validity: None given
Scoring: Most subtest scores convert to grade level scores. Each subtest offers an extensive Checklist of Reading Difficulties.
Limitations: Since it lacks information on norms, reliability, and validity, the norms could be deceiving. The examiner must focus on the qualitative information offered by the test.

GATES-McKILLOP READING DIAGNOSTIC TEST

Arthur I. Gates and Anne S. McKillop
Publisher: Teachers College Press
Date: 1962
Test Objectives: To assess skill development in reading of children grades 2-6 through a battery of 11 subtests.
Subtests:
1. Oral Reading - Consists of a series of graded reading passages. The student reads orally while the examiner records errors and notes reading characteristics on a record booklet.
2. Words: Flash Presentation - Uses a tachistoscope to show words in one-half second.
3. Words: Untimed Presentation - The student reads words without time restriction.
4. Phrases: Flash Presentation - The student reads phrases shown in the tachistoscope for one-half second.
5. Knowledge of Word Parts: Word Attack - This subtest has four parts: Recognizing and Blending Common Word Parts, Giving Letter Sounds, Naming Capital Letters, Naming Lower-Case Letters.
6. Recognizing the Visual Form or Word Equivalents of Sounds - This subtest has four parts: Nonsense Words, Initial Letters, Final Letters, Vowels.
7. Auditory Blending - The examiner pronounces words in parts and the student blends the parts to form the word orally.
8. Spelling - The examiner reads words aloud while the student writes them.
9. Oral Vocabulary - The student must define words.
10. Syllabication - The student divides words into syllables.
11. Auditory Discrimination - The student discriminates among common English phonemes.

Norm Group: The manual gives many norm tables but no information about the norm population.
Reliability: None given
Validity: None given
Scoring: The test has two conversion tables, grade score and interpretation. The interpretation tables allow the examiner to compare performance on the Oral Reading subtest with each of the subtests assessing specific skills.
Limitation: Information regarding the norm population, reliability, and validity is absent. Scoring and interpretation is complex and difficult.

GILMORE ORAL READING TEST

Authors: John V. Gilmore and Eunice C. Gilmore
Publisher: Harcourt, Brace, Jovanovich
Date: 1968
Test Objectives: This test is designed to test oral reading in grades 1 to 8.
Subtests:
1. Oral Reading - Contains ten paragraphs of increasing difficulty.
2. Comprehension - The examiner records answers to five comprehension questions.
3. Rate - The passages are timed but the rate is scored separately.

Norm Group: The authors used 4,455 pupils in six communities. They represented a variety of socio-economic backgrounds.
Reliability: (alternate form)

Grade	Number	Accuracy	Comprehension	Rate
3	51	.94	.60	.70
6	55	.84	.53	.54

Validity: No validity established for current test. The manual gives validity of earlier test.
Scoring: This test provides grade scores and performance ratings for accuracy and comprehension. Rate of reading is scored as slow, average, or fast.
Limitations: Technical data is limited.

GRAY ORAL READING TEST

Author: William S. Gray
Publisher: Bobbs-Merrill Co., Inc.
Date: 1963
Test Objectives: To test skill development in oral reading preprimer to adult.
Subtests:
1. Oral Reading - Contains 13 passages of increasing difficulty.
2. Comprehension - The examiner asks four recall questions.
3. Rate - The readings are timed and that time enters into the overall passage score.

Norm Group: The norms are limited. They are based on only 502 children in Florida, Chicago and its suburbs.
Reliability: (alternate form) - The data consists of intercorrelations among grade scores on each of the four forms.
.97 - .98 for girls
.96 - .98 for boys
Validity: The author bases test validity on the procedures used in its construction and on the discrimination between students at different grade levels.
Scores: This test gives a grade score for speed and accuracy of oral reading. It also gives error analysis.
Limitations: The technical data is limited. Recording and scoring is difficult although a tape recorder helps.

PEABODY INDIVIDUAL ACHIEVEMENT TEST

Authors: Lloyd M. Dunn, Ph.D. and F. C. Markwardt, Jr., Ph.D.
Publisher: American Guidance Service
Date: 1970
Test Objectives: To test a wide-range of academic achievement in five content areas which includes two in reading. It is designed to be used with students in kindergarten through eighth grade.
Subtests:
1. Reading Recognition - This contains eighty-four items used to assess skill development in matching letters, naming capital and lower-case letters, and recognizing words in isolation.
2. Reading Comprehension - This contains sixty-six multiple choice items to determine comprehension. The student demonstrated this by choosing the appropriate picture out of four.

Norm Group: A total of 2,899 children were used. The author used percentages from the 1967 census as a guide for selection of the standardization sample.
Reliability: (test-retest) - Median reliabilities are .64 for Reading Comprehension and .89 for Reading Recognition.
Validity: The authors justify the content validity by the use of expert opinion and extensive reviews of curriculum material. Vocabulary Test and the Reading Recognition subtest average .541, Comprehension was .66.
Scoring: The raw scores can be translated into four derived scores; grade equivalents, age equivalents, percentile ranks, and standard scores.
Limitations: The reliability and validity are questionable.

SILVAROLI CLASSROOM READING INVENTORY

Author: Nicholas J. Silvaroli
Publisher: William C. Brown Book Co.
Date: 1974
Test Objectives: Silvaroli designed this test as a replacement for the informal inventory. It can be used with students from preprimer to grade 6.
Subtests:
1. Graded Word List - The student pronounces words beginning at preprimer and stopping where he fails on five or more words.
2. Graded Oral Paragraphs - This begins at the highest word list level. The examiner records oral errors.
3. Spelling Survey - This includes fifteen words from each level.

Norm Group: None given
Reliability: None given
Validity: None given
Scoring: Scores are given as functional and instructional. The latter requires 95% correct word recognition and 75% comprehension. Each selection has five comprehension questions which include factual, inference, and vocabulary items. The examiner notes any consistent errors in any one type of question.
Limitations: The technical data is absent and the grade 6 ceiling is low.

STANFORD DIAGNOSTIC READING TEST

Authors: Bjorn Karlson, Richard Madden, and Eric Gardner
Publisher: Harcourt, Brace, Jovanovich
Date: 1976
Test Objectives: Designed to test a series of specific reading skills. There are four overlapping levels of the test which are color coded.
Subtests:
1. Red Level - grades 1.5 to 3.5, 6 scores; Auditory Discrimination, Phonetic Analysis, Auditory Vocabulary, Word Reading, Reading Comprehension (2 parts); 150 min., 3-5 sessions.
2. Green Level - grades 2.5, 5.5, 7 scores; Auditory Discrimination, Phonetic Analysis, Structural Analysis, Auditory Vocabulary, Reading Comprehension (2 parts); 113 min., 3-5 sessions.
3. Brown Level - grades 4.5 to 9.5, 7 scores; Phonetic Analysis, Structural Analysis, Auditory Vocabulary, Reading Comprehension (2 parts), Reading Rate; 113 min., 1-5 sessions.
4. Blue Level - grades 9-13, 12 scores; Phonetic Analysis, Structural Analysis, Word Meaning, Word Parts, Reading Comprehension (2 parts), Fast Reading, Scanning and Skimming.

Norm Group: The authors used data from the United States Office of Education's 1970 census to create a random sampling of 3,000 school districts.
Reliability: The internal-consistency coefficients range from .85 to .90. Alternate form coefficients range from .75 to .94.
Validity: The content validity is based on the extent to which the test content reflects local curriculum. The criterion-related validity is .61 to .98 for the Red and Green levels and .39 to .94 for the Brown level.
Scoring: This test is norm-referenced and criterion-referenced. It can be machine or hand scored. The raw scores can be transformed into "Progress Indicators" (the only criterion-referenced score), percentile scores, grade equivalents, and scaled scores.
Limitations: Children must be able to read and write. The test is too long.

WIDE RANGE ACHIEVEMENT TEST

Authors: J. F. Jastak and S. R. Jastak
Publisher: Guidance Associates
Date: 1965
Test Objectives: This is a test to assess reading, spelling, and math using two levels. Level I is for students younger than 12 and Level II is for those over 12.
Subtests: 1. Reading - This subtest assesses skills in recognizing capital letters, naming capital letters, and recognizing words in isolation.
Norm Group: Level I used 5,868 people, Level II used 5,933 for the standardization process. No attempt was made to find a representative national sample.
Reliability: The authors report that split-half reliabilities exceed .90.
Validity: The WRAT has a .91 correlation with the Gilmore Oral Reading Test accuracy score.
Scoring: Each subtest uses grade equivalent, percentile ranks within grades and standard scores.
Limitations: The WRAT has limited norms and limited behavior sampling.

READING READINESS TESTS

Reading readiness tests are given to a majority of students entering school at the first grade level throughout the United States. Their popularity alone suggests that they are of some use to teachers in making decisions about children. However, the entire concept of "readiness" has recently been under attack, which makes these tests somewhat controversial. The following material dealing with readiness tests will be presented in three sections. First, there will be a general description of "readiness" and how to assess it. Second, the particular technical problems that face readiness tests will be discussed, along with the uses (or purposes) of a readiness test. The third section will contain brief descriptions of four tests designed to measure school (reading) readiness. (Salvia, Ysseldyke, 1978).

Readiness can be defined as being a collection of factors, resulting in a state of general maturity, which allow a child to read without excess difficulty. Readiness as it applies for the first grade or kindergarten, refers to both academic and social readiness. The major characteristics that are important in reading readiness are chronological age, sex, mental age or general intelligence, language development, and both social and emotional maturity. (Harris & Sipay, 1975) Formal readiness assessment usually deals with academic process testing. Therefore, intelligence, or aptitude, which underlies all school achievement, is often a major component

of a readiness test. Perceptual-motor development and lan-uage development (although these are less predictive of future school achievement) are also important and therefore measured by readiness tests. (Salvia, Ysseldyke, 1978)

There are three distinct features that distinguish a readiness test:

1. Readiness tests are usually administered during kindergarten or upon entry into first grade.
2. These tests are used to predict initial school success, and isolate those who may need remedial programs or delayed school entry.
3. These tests usually contain the word readiness somewhere in the test name. (Salvia, Ysseldyke, 1978)

The purposes of a reading readiness test are twofold and correspond with the second feature of a readiness test. They are: 1) To predict who is not yet ready for formal instruction. 2) To predict those who will benefit from additional programs. It should be noted that a major purpose of these tests is NOT to predict achievement, but rather to select those with difficulties.

Validity, especially predictive validity, is a major consideration in selecting a readiness test. It must prove (through data in the technical manual) to the user that a child with poor scores on the readiness test will perform poorly during actual schooling. Predictive validity should

also enable the user to determine the importance of various readiness skills. (Farr & Anastasiow, 1975)

Further evaluation of a reading readiness test should include: 1) A careful examination of test procedures and test formats. Young examinees may be penalized by unusual formats or procedures. (Farr & Anastasiow, 1978) 2) The reliability coefficient should be at least .90 for satisfactory measurement of an individual student. (Sylvia, Ysseldyke, 1976) 3) The norms should be recent, representative, and relevant to the school population. 4) The test should be relatively easy to administer. Parts which are individually administered are very time consuming for the classroom teacher. 5) The test should be easy to score and interpret. It should also provide an insight into a student's strengths and weaknesses. The user should be wary of a large number of very brief subtests, as they may be unreliable when considered separately. 6) There should be a positive correlation (.5 or greater) between total reading readiness scores and reading scores at the end of first grade. (Harris & Sipay, 1975) Teachers should also be aware that scores obtained from preschool students are often highly unreliable. They are often subject to rapid change based upon the effectiveness of instruction and various developmental changes. (Farr & Anastasiow, 1975, p. 31) It cannot be emphasized too strongly that these readiness tests should not be the sole criteria for evaluating a pupil's readiness. Teacher judgment also plays an equally important role.

A brief overview of four of the many reading readiness tests available follows. A chart is placed at the end of the descriptions for quick reference.

LEE-CLARK READING READINESS TEST (LCRRT)

Authors: J. Murray, Lee and Willis Clark
Publisher: CTB/McGraw-Hill
Date: 1962
Range: Kindergarten - 1
Levels: 1
Time: 20 minutes - 2 sessions
Subtests: Letter matching, letter discrimination, picture vocabulary, letter and word recognition. A total score is also provided.
Norms: Are of unknown appropriateness. The LCRRT was standardized on over 1,000 children of unknown age, sex, or demographic characteristics.
Reliability: Split-half coefficients based on the total raw score range from .80 to .96. Subtest reliability is low for diagnostic use (.56 to .88). (Salvia, Ysseldyke, 1978)
Predictive Validity: Predictive validity coefficients range from .13 to .72 (median .54) when compared with teacher ratings and .14 to .79 (median .45) when compared with scores on the Lee-Clark Reading Test: First Readers. This data indicates that the LCRRT lacks the predictive validity necessary to be the only instrument used in making educational decisions.
Manual: Good. Directions are clearly and exactly stated.
Summary: The Lee-Clark Reading Readiness Test is one of the better known readiness tests. It is a group administered, norm referenced, multiple skill device designed to predict which children are ready to begin reading instruction and which children need a "special developmental program." (Salvia, Ysseldyke, 1978, p. 391) This test is designed for use with children finishing kindergarten or entering first grade. It is easy to administer, has a short testing time, and is easily interpreted. However, it has questionable predictive validity and inadequately described norms which affect its usefulness. This entire test may be used as a gross screening device, but should not be used to provide diagnostic assessment of strengths and weaknesses.

METROPOLITAN READINESS TEST

Authors: Joanne Nurss, Mary McGauvran
Publisher: Harcourt, Brace, Jovanovich
Date: 1976
Range: Kindergarten - 1
Levels: 2 - Level 1 - beginning or middle kindergarten
Level 2 - end of kindergarten, beginning 1st grade
Time: Untimed. Usually requires 80-90 minutes over several sessions.
Subjects: Level 1 - auditory memory, rhyming, letter recognition, visual matching, school language and listening, quantitative language (optional - copying).
Level II - beginning consonants, sound-letter correspondence, visual matching, finding patterns, school language, listening (optional - quantitative concepts and operations).
Norms: Adequately described.
Reliability: Total scores are sufficiently reliable to make instructional decisions (.85 to .95), while subtests are not (.67 to .90).
Predictive Validity: Established by correlating the Metropolitan Readiness Scores with achievement test scores on the Metropolitan Achievement Tests. Correlation was .70.
Summary: The Metropolitan Readiness Test is a norm referenced, group administered multiple skill battery designed to assess the skills necessary for early success in school. This test has a very complete manual, and the authors considered the age of the examinees in the format and test procedures. A practice test is provided, test taking concepts and skills are taught, and directions are clear. This instrument appears to be adequately normed, has adequate reliability, and adequate validity. The authors have attempted to eliminate any cultural or sex bias. The time to administer the entire test may be a drawback in some classroom situations. This test can provide useful information in screening first graders.

MURPHY-DURRELL READING READINESS ANALYSIS

Publisher: Harcourt, Brace, Jovanovich
Date: 1965
Range: K.8 0 1.5
Levels: 1
Time: 60 minutes - 2 sessions
Subtests: Phonemes, Letter names, Learning rate
Manual: Good
Norms: Inadequately described. No information of pertinent data is included in the manual. Regional distribution is quite uneven. Test results may be more interpretable if local norms are developed. No attempt at minority representation is given.
Reliability: The manual lists reliability on the total score of .98, while subtests range from .88 to .97.
Predictive Validity: Correlation with Stanford Achievement Test at the end of first grade listed at .65. However, this evidence was based on a small sample of 200 pupils in Kansas and limits the interpretation of this validity data.
Summary: The Murphy-Durrell Reading Readiness Analysis is a norm-referenced, group administered test, designed to measure the abilities essential for success in beginning reading. All of the subtests are closely related to the actual task of learning to read. However, the skills are primarily related to the decoding aspects of beginning reading, and other factors should be considered if this test is to be used as an indication of readiness. The Learning Rate Subtest is a measure of the number of sight words retained (learned) after instruction. There is no pretest, so the possibility exists that prior knowledge may be tested - not new learning. There are suggestions for grouping based on test results. The use of this test is probably most useful as a measure of a specific skill, and should be used with local norms. (Farr & Anastasiow, 1975)

THE MACMILLAN READING READINESS TEST, Revised Edition

Authors: Albert Harris, Edward Sipay
Publisher: Macmillan Company
Date: 1970
Range: Kindergarten - 1
Levels: 1
Time: 75-90 minutes, 3-5 sessions
Manual: Difficult to use. Instructions are ambiguous.
Subtests: Visual discrimination, auditory discrimination, vocabulary and concepts, letter names, visual motor, and a rating scale.
Norms: Adequately described, although no provision for minorities was made. A review of this test suggests its use with middle-class children, and points out that vocabulary items may be biased in their favor. (Buros, 1978)
Reliability: Satisfactorily high (.80 to .95) for total score.
Predictive Validity: A correlation of .66 was given in predicting first grade success in reading. The authors invite the test user to judge the content validity of this test for themselves. (Buros, 1978)
Summary: The Macmillan Reading Readiness Test, Revised Edition, is a norm referenced, group administered test of reading readiness designed to aid in "determining which children are ready for beginning reading instruction." (Buros, 1978, p. 1338) The usefulness of this test lies in helping a teacher decide what a student knows, or not, and then use this information in a plan for instruction. This test, however, requires a "Super Teacher" to correctly administer and interpret the results. If directions in the manual are followed, it would take five hours of testing time, and ten hours to score - with the visual motor subtest being most difficult. Also, the manual is hard to follow, and instructions are confusing. The teacher would also have difficulty scoring the rating scale. He/she must decide which of five ratings to assign to a student's pronunciation, muscular coordination, and a variety of experiences. It has been suggested that this test may be more appropriate for a phonetic approach to reading. (Buros, 1978)

Reading Readiness Test	Date of Publication	Range	Levels	Time/ Sessions	Reliability	Norms	Predictive Validity	Manual	Subtests
Lee-Clark Reading Readiness Test	1962	K-1	1	20 min.2 session	Adequate for total score	Not described adequately	Inadequate	Adequate	Letter matching Letter discrimination-picture voc. letter & word rec.
Metropolitan	1972	K-1	2	Untimed/ several	Adeqaute for total score	Adequate	Adequate	Good	Level I - 8 Level II - 8
Murphy-Durrell Reading Readiness Analysis	1965	.8-1.5	1	60 min/2 sessions	Adequate for total score	Not described adequately	Inadequate	Adequate	Phonemes Letter Names Letter Rate
Macmillan Reading Readiness Test, Revised Edition	1970	K-1	1	75-90/ 3 to 5 sessions	Adequate for total score	Adequate	Adequate	Difficult to use	Visual & auditory discrimination, voc. & concepts letter names visual motor + a rating scale

CRITERION-REFERENCED TESTS

A criterion-referenced test relates test performance to absolute standards. It indicates whether the student has met specific criteria, rather than how well his performance compares with that of others. A main advantage claimed for criterion-referenced tests is their utility for improving educational programs - since they provide specific information on what the student knows and does not know, the tests are useful in individualizing instruction. Other advantages are their flexibility, their adaptability for continuous assessment, and the fact that the student is not labeled with some statistic in comparison with other students. Disadvantages are the setting of inappropriate criteria and what to do if the child does not learn the task.

There are several categories of criterion-referenced tests. There is the objectives-based test in which every item is directly identified with an explicitly stated educational behavioral objective. The objectives may be broad in scope, so that the test serves as an achievement or survey test. The objectives may also be very specific, so as to serve as a diagnostic test. Several advantages have been cited for these objectives-based tests, the main one being that they can be tailor-made to assess mastery of specific objectives in a curriculum, and therefore useful in individualizing instruction. Also, because of the stated objectives, the examiner

ought to know exactly what is being tested. Some disadvantages include the hierarchy of skills that is set forth in these tests when there is no basis for this sequence. Mastery levels are arbitrarily set. (NEA, 1977) Unless the objectives are good, the test is of no use. Test items frequently represent skills in too isolated a context, rather than in an applied or functional situation. (Schreiner, 1979)

Another category of the criterion-referenced test is the informal test. This includes the Informal Reading Inventory (IRI) and Cloze tests. Advantages of the informal test are the simple construction, the child is apt to be more comfortable with the informal procedure, they are inexpensive in time and money, they can be closely related to instruction, and information obtained can be readily applied. The IRI consists of a series of graded passages for oral and silent reading. Three levels of reading are determined by set criteria. They are:

Independent - 99% word recognition and 90% comprehension

Instructional - 95% word recognition and 75% comprehension

Frustrational - below 90% word recognition and 50% comprehension

There are problems concerning the reliability and adequacy of comprehension questions on the IRI's, but the content validity is usually sufficient for forming instructional reading groups. (Schreiner, 1979) Many basal reading series include an IRI constructed from their readers.

The Cloze test consists of a passage with every Nth word (usually every fifth word) deleted. Students are then to read the material and attempt to fill in the blanks, using the correct word according to the proper context of the sentence. The percentage of correct answers is then calculated. The criteria for determining levels for instruction is: Independent Level - 60% or better; Instructional Level - 35%-59%; Frustration Level - 34% or below. Rankin and Culhane (1969) studied the validity of the cloze procedure and stated "the cloze procedure is a highly valid measure of reading comprehension. The average validity coefficient was .68. An advantage of this test over the IRI is that it can be administered to a group. (Ekwall, 1976) The teacher must be aware that these procedures do not consider the interest level of the material, or the concept load of the material. These two factors can result in different comprehension scores.

Guidelines for selecting published criterion-referenced tests are much the same as for norm-referenced tests. Reliability should be considered. Validity, especially content validity, is important. Unfortunately, most published criterion referenced tests at this time are lacking this information. Other factors to consider are the objectives -- Are they stated clearly? Do they match those of the school program? The test items -- Are the items valid indicators of the objectives they were developed to measure? Are the items free of technical flaws? Have the test items been field tested?

SUMMARY

Although this chapter has dealt mainly with testing instruments designed to measure reading performance objectively, it must be noted that these instruments represent only a portion of the total evaluation process.

In reading, assessment begins when a student's performance is measured against the goals and objectives that were set for instruction. (Schreiner, 1979) This measuring (or testing) of performance must be comprehensive, continuous, and articulated throughout the school system, as well as being functional and practical. To be functional and practical, the testing program should give the specific information desired by both teachers and administrators, and should use tests that are easy to give, score, and interpret. All-school planning is essential to achieve continuity and articulation throughout the grades. (Salvia, Ysseldyke, 1978)

A comprehensive testing program should include: (1) teacher observation, (2) health and anecdotal records, (3) work samples, (4) classroom tests, (5) cumulative records, (6) case studies by specialized consultants, (7) objective or standardized tests, and (8) any other tests needed to assess student performance. (Harris & Sipay, 1975) Reading tests are an integral part of this needs assessment program.

BIBLIOGRAPHY

1. Austin, Mary C.; Bush, Clifford; and Huebner, Mildred. Reading evaluation. New York: Ronald Press, 1961.

2. Baumann, James, and Stevenson, Jennifer. Standardized reading achievement tests: potential uses and abuses. n.p., n.d.

3. Berg, Paul Conrad. Evaluating Reading Abilities, in Assessment in Reading Problems. Newark, Delaware: International Reading Association, 1973.

4. Blanton, William; Farr, R.; and Tuinman, J. eds. Measuring reading performance. Newark, Delaware: International Reading Association, 1974.

5. Buros, Oscar K. ed. Tests in print II. Highland Park, N. J.: Gryphon Press, 1974.

6. Buros, Oscar K. ed. The eighth mental measurements yearbook. Vol. II. Highland Park, N. J.: Gryphon, Press, 1978.

7. Buros, Oscar K. ed. The seventh mental measurements yearbook. Vol. II. Highland Park, N. J.: Gryphon, Press, 1974.

8. California Testing Bureau. A glossary of measurement terms: a basic vocabulary for evaluation and testing. Monterey, California: McGraw-Hill, 1973.

9. Ekwall, Eldon E. Diagnosis and remediation of the disabled reader. Boston: Allyn and Bacon, 1975.

10. Farr, Roger, and Anastasiow, Nicholas. Tests of reading readiness and achievement: a review and evaluation. Newark, Delaware: International Reading Association, 1975.

11. Gronlund, Norman E. Measurement and evaluation in reading. New York: Macmillan, 1976.

12. Hambleton, Ronald, and Eignor, Daniel. Guidelines for evaluating criterion-referenced tests and test manuals. Journal of Educational Measurement, 1978, 15.

13. Harris, Albert, and Sipay, Edward. How to increase reading ability (6th ed.). New York: Longman, 1975.

14. Jongsma, Eugene A. Gates-MacGinitie reading tests. Journal of Reading, 1980, 4.

15. Katz, Martin. Selecting an achievement test: principles and procedures. Princeton, N. J.: Educational Testing System, 1973.

16. MacGinitie, Walter, ed., Assessment problems in reading. Newark, Delaware: International Reading Association, 1973.

17. MacGinitie, Walter H., What are we testing? in Assessment in reading problems. Newark, Delaware: International Reading Association, 1973.

18. Martuza, Victon R. Applying norm-referenced and criterion-referenced measurement in education. Boston: Allyn and Bacon, 1977.

19. National Education Association. Standardized testing issues: teacher's perspectives. Washington, D. C.: NEA, 1977.

20. Russell Sage Foundation. Teachers and testing. New York, 1967.

21. Salvia, John and Ysseldyke, James. Assessment in special and remedial education. Boston: Houghton Mifflin, 1978.

22. Schreiner, Robert, ed. Reading tests and teachers: a practical guide. Newark, Delaware: International Reading Association, 1979.

23. Spache, George D. Diagnosing and correcting reading disabilities. Boston: Allyn and Bacon, 1976.

24. Spache, George D. Investigating the issues of reading disabilities. Boston: Allyn and Bacon, 1976.

25. Wallace, G., and Larsen, S. C. The educational assessment of learning problems: testing for teaching. Boston: Allyn and Bacon, 1978.

26. Wick, J. W. Educational measurement: where are we going and how will we know when we get there? Columbus, Ohio: Merrill, 1973.